

Noise Resilient Recovery Algorithm for Compressed Sensing

V. Meena and G. Abhilash
meena_pec10@nitc.ac.in abhilash@nitc.ac.in

Department of Electronics and Communication Engineering
National Institute of Technology Calicut
Kerala, 673 601 India

Abstract

In this article, we discuss a novel greedy algorithm for the recovery of compressive sampled signals under noisy conditions. Most of the greedy recovery algorithms proposed in the literature require sparsity of the signal to be known or they estimate sparsity, for a known representation basis, from the number of measurements. These algorithms recover signals when noise level is significantly low. We propose Entropy minimization based Matching Pursuit (EMP) which has the capability to reject noise even when noise level is comparable to that of signal level. The proposed algorithm can cater to compressible signals and signals for which sparsity is not known in advance. Simulation study of the proposed scheme shows improved robustness to white Gaussian noise in comparison with the conventional greedy recovery algorithms.

Keywords: Sparse representation, measurement matrix, entropy based matching pursuit, greedy recovery algorithms, compressed sensing

1 Introduction

Compressed sensing (CS) is a signal acquisition scheme that embeds the intelligence of compression along with signal acquisition in discrete form. The utility of CS is more pronounced in acquiring wide-band signals which have a sparse representation in some domain.

If $\{\Psi_i(t); i = 1, 2, \dots, N\}$ is a representation basis for $\mathbf{S}(t) \in \mathbb{R}^N$, then $\mathbf{S}(t) = \sum_{i=1}^N C_i \psi_i(t)$, where $C_i, i = 1, 2, \dots, N$ are representation coefficients, which form an $N \times 1$ vector $\mathbf{C} = [C_1, C_2, \dots, C_N]^T$. For signals in \mathbb{R}^N , the ℓ_0 norm is defined as

$$\|\mathbf{C}\|_0 = |\text{supp}(\mathbf{C})| = |\{i : C_i \neq 0\}|. \quad (1)$$

$\mathbf{S}(t)$ is K -sparse if $\|\mathbf{C}\|_0 \leq K$ with $K \ll N$. Real signals are rarely sparse, but are compressible. A signal is compressible if it can be represented in an appropriate basis with only a few significant coefficients, *i.e.*; when sorted in the descending order, the coefficients follow power-law decay as,

$$|C_i| \leq P i^{-r}, \text{ for } i = 1, 2, 3, \dots \text{ and } r > 1, \quad (2)$$

where the non-negative constant P is independent of r . The lower bound on the number of measurements required for stable recovery is a function of the sparsity K of the signal. The basic requirement in CS is to identify a representation basis relative to which sparsity K is minimum.

A measurement matrix which is incoherent or least correlated with Ψ measures the signal at a rate much lower than the Nyquist rate required for the signal. If the number of measurements is greater than its sparsity, it would be possible to recover the signal in a stable manner [1, 2]. To ensure that the

geometry of sparse signals is preserved in the measurements, the matrix Φ of measurement functions should satisfy the Restricted Isometry Property (RIP) [3][1]:

$$(1 - \delta_K) \|\mathbf{S}(t)\|^2 \leq \|\Phi \mathbf{S}(t)\|^2 \leq (1 + \delta_K) \|\mathbf{S}(t)\|^2; \quad 0 < \delta_K < 1, \quad (3)$$

where δ_K is the restricted isometry constant (RIC) corresponding to the K sparse signal $\mathbf{S}(t)$. The RIP requires the energy of the measurement to remain within a closed limit around the energy of the signal. The recovery of the sparse set of representation coefficients becomes more stable as δ_K approaches zero. The measurement vector $\mathbf{y} = [y_1, y_2, \dots, y_M]$ can be obtained as the inner products of the signal $\mathbf{S}(t) \in \mathbb{R}^N$ with the M measurement functions $\phi_i(t)$ for $i = 1, 2, \dots, M$ evaluated over a finite duration. That is, $y_i = \langle \mathbf{S}(t), \phi_i(t) \rangle$; for $i = 1, 2, \dots, M$. The measurement vector is

$$\mathbf{y} = \Phi \Psi \mathbf{C} = \mathbf{A} \mathbf{C} = \Phi \mathbf{S}(t), \quad (4)$$

where $\mathbf{A} = \Phi \Psi$, with Φ and Ψ having sizes $M \times N$ and $N \times N$, respectively. To achieve stable recovery of the sparse set of coefficients and reconstruction of the signal, the measurement functions should be chosen such that they capture maximum information of $\mathbf{S}(t)$. In general, $K \leq M$ and $M \ll N$. For convenience, in the remaining part of the article, we use \mathbf{S} in place of $\mathbf{S}(t)$.

To minimize the number of measurements M , the measurement functions $\{\phi_i(t)\}_{i=1}^M$ should not be able to sparsely represent the representation basis $\{\psi_j(t)\}_{j=1}^N$ relative to which the signal is sparse, and vice versa [2]. Thus,

the mutual coherence between the two matrices Ψ and Φ denoted by

$$\mu(\Phi, \Psi) = \sqrt{N} \max\{|\langle \psi_j, \phi_i \rangle|, \text{ for } 1 \leq i \leq M, 1 \leq j \leq N\} \quad (5)$$

should be minimum. For normalized matrices, μ is within $[1, \sqrt{N}]$ [1]. Thus, for compressed sensing of signals, a representation basis in which signal is exactly sparse or is compressible is to be identified. A measurement matrix, which is incoherent with the representation basis, and algorithms necessary to recover the sparse representation coefficients from the compressed measurements constitute the other vital components in CS.

In [4], Peyre proposes a method to reconstruct the signal from \mathbf{y} when information about the measurement matrix alone is known. In this method, estimation of representation basis and recovery of sparse signal are done simultaneously. The representation basis which results in maximum sparsity of the signal is estimated from a tree structured dictionary of orthogonal bases using iterative thresholding algorithm. In [5], Ravishankar and Bresler discuss a method for learning sparsifying transform from the data. They propose a generalized formulation of transform learning at the analysis side that learns well-conditioned transforms under both noiseless and noisy conditions.

The commonly used iterative recovery algorithms assume knowledge of the representation basis, in which the signal is sparse. These algorithms find an approximation to the signal by minimizing the residual energy [3][6]-[11] under high signal to noise ratio. Greedy pursuit algorithms like Orthogonal Matching Pursuit (OMP) [7], its variants generalized OMP (gOMP) [8], Regularized OMP (ROMP) [12] and Compressed Sampling Matching Pursuit (CoSaMP) [10] exhibit good performance when sparsity is known in

advance. These algorithms use sparsity as a parameter. They also consider unrecoverable energy to be greater than noise and therefore require noise level to be significantly low compared to signal level. In [13], compressed sensing of a signal of interest corrupted by an interfering signal is filtered to separate the signal of interest from noise. But orthogonality condition is imposed on the noise subspace with respect to the signal subspace for achieving the desired result. In practical scenario, it is not easy to meet these constraints. These constraints can be removed if we resort to methods which do not consider ℓ_2 norm directly for choosing the support. In [14], best representation basis is identified adaptively, from a dictionary of wavelet packets by choosing the decomposition structure which minimizes Shannon entropy.

The entropy minimization based matching pursuit (EMP) algorithm proposed in this article is motivated by the fact that sparsity can be induced by minimizing Shannon entropy of signal representation. In the sequel, entropy means Shannon entropy. The advantage offered by EMP algorithm is its noise resilience during signal recovery. In the absence of noise, the performance of EMP algorithm is at par with the Matching Pursuit (MP) algorithm in terms of Signal to Reconstruction Error Ratio (SRER) [15]. The EMP algorithm can be used to arrive at a sparse representation of a signal when its representation basis is known. Sparse representation and signal recovery are considered dual to each other in [7].

In this article, EMP algorithm is used for signal recovery from measurements in the context of compressed sensing. Compared to the conventional greedy pursuit algorithms, the EMP algorithm has superior capability to extract signal components from noisy measurements. In Section 2, we focus on the formulation of the EMP algorithm along with its performance

analysis and proof of convergence. Section 3 presents results of simulation study carried out on synthetic sparse signals and a class of signals for which sparsity is not known upfront. The class of signals chosen is speech signals. Results are presented for both noise-free and noisy cases. A discussion on the results is also presented. The article is concluded in Section 4.

2 Entropy minimization based Matching Pursuit Algorithm for signal recovery

Matching pursuit (MP) algorithm, OMP algorithm, ROMP algorithm and CoSaMP algorithm are examples of greedy iterative pursuit algorithms. Each iteration updates $\hat{\mathbf{y}}$, the approximation of \mathbf{y} in (4), such that the residual energy $\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2$ is minimized. The update $\hat{\mathbf{y}}$ is obtained by choosing one or more columns from the matrix \mathbf{A} , that correlated best with the residual error vector resulting from the previous iteration.

The EMP algorithm is a variant of MP Algorithm. It minimizes the overall entropy of the signal representation in each iteration instead of minimizing the residual energy. Entropy $H(\mathbf{S})$ of the representation of the signal \mathbf{S} is related to the theoretical dimension N of the signal as [14, 16]

$$N = \exp(H(\mathbf{S})). \quad (6)$$

In [15], the EMP algorithm is used for obtaining a sparse representation of a class of signals from its representation in time domain assuming that the sparsifying frame is known. In the context of compressed sensing, we extend the algorithm for recovering the sparse representation \mathbf{C} of \mathbf{S} from the measurements \mathbf{y} . The estimate $\hat{\mathbf{C}}$ of \mathbf{C} , thus obtained, is used to obtain an estimate $\hat{\mathbf{S}}$ of the signal \mathbf{S} through $\hat{\mathbf{S}} = \Psi\hat{\mathbf{C}}$.

Without loss of generality, we consider a normalized signal $\mathbf{X} = \{x_i, i = 1, 2, \dots, N\}$. x_i^2 represents the probability of choosing the i -th function of some basis in \mathbb{R}^N . The entropy of representation of \mathbf{X} is defined by

$$H(\mathbf{X}) = \sum_{i=1}^N x_i^2 \log \frac{1}{x_i^2}. \quad (7)$$

Let the entropy of the representation of the signal \mathbf{y} be denoted as $H(\mathbf{y})$. At iteration m , the conditional entropy of the representation of \mathbf{y} , given the vectors in the set $\hat{\mathbf{A}}^{(m)}$, is represented as $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$, where $\hat{\mathbf{A}}^{(m)}$ is the matrix which approximates \mathbf{A} at the m^{th} iteration. Similarly, let the conditional entropy of representation of the residual signal \mathbf{e} , given \mathbf{y} and $\hat{\mathbf{y}}$, be $H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)})$ and the conditional entropy of the representation of $\hat{\mathbf{y}}$, given the vectors in the set $\hat{\mathbf{A}}^{(m)}$, be $H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)})$. The mutual information between $\hat{\mathbf{y}}$ and $\hat{\mathbf{A}}^{(m)}$ is denoted as $I(\hat{\mathbf{y}}, \hat{\mathbf{A}}^{(m)})$. In every iteration, signal \mathbf{y} is represented as the sum of approximation $\hat{\mathbf{y}}$ and residual \mathbf{e} with the available \mathbf{A} . Aim is to choose the smallest subset of \mathbf{A} to estimate $\hat{\mathbf{y}}$.

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e} \quad (8)$$

Though \mathbf{e} is completely determined by \mathbf{y} and $\hat{\mathbf{y}}$, sparsity of \mathbf{e} or representation entropy of \mathbf{e} will be determined by the choice of $\hat{\mathbf{A}}^{(m)}$ which determines $\hat{\mathbf{y}}$.

$$H(\mathbf{y}|\hat{\mathbf{A}}^{(m)}) = H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)}) + H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)}) \quad (9)$$

Using the definition of mutual information,

$$H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)}) = H(\hat{\mathbf{y}}) - I(\hat{\mathbf{y}}, \hat{\mathbf{A}}^{(m)}) \quad (10)$$

Since $\hat{\mathbf{y}} = \hat{\mathbf{A}}^{(m)}\hat{\mathbf{C}}$, $H(\hat{\mathbf{C}})$ is the information left in $\hat{\mathbf{y}}$ when prior information about $\hat{\mathbf{A}}^{(m)}$ is available. Thus,

$$H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)}) = H(\hat{\mathbf{C}}) \quad (11)$$

Substituting (11) in (9),

$$H(\mathbf{y}|\hat{\mathbf{A}}^{(m)}) = H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)}) + H(\hat{\mathbf{C}}). \quad (12)$$

$H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)})$ is the entropy of representation of residual error. A sparse $\hat{\mathbf{C}}$ results by minimizing the conditional entropy $H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)})$. On convergence of the algorithm, $H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)})$ is zero or is negligible compared to $H(\hat{\mathbf{C}})$. Thus, minimization of $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$ leads to minimum $H(\hat{\mathbf{C}})$. By (6), our aim is to minimize $H(\mathbf{S})$ in order to reduce the dimension of \mathbf{S} which is achieved by minimizing $H(\mathbf{y}|\mathbf{A})$. This is attained by solving

$$\min_{\mathbf{C}} \{H(\mathbf{C})\} \text{ subject to } \mathbf{y} = \mathbf{AC}; \mathbf{A} = \mathbf{\Phi\Psi}. \quad (13)$$

Since the signal can be normalized, $0 \leq e_i^2 \leq 1$, where e_i is the i^{th} component of \mathbf{e} , and e_i^2 can be considered as the probability of occurrence of the component e_i . Thus entropy of the representation of residue conditioned on the estimated $\hat{\mathbf{y}}$ is calculated as $\sum_{i=1}^M e_i^2 \log \frac{1}{e_i^2}$. Similarly, $H(\hat{\mathbf{C}})$ is calculated using the normalized vector of sparse measurements as $\sum_{i=1}^N \hat{c}_i^2 \log \frac{1}{\hat{c}_i^2}$, where \hat{c}_i is the i^{th} component of $\hat{\mathbf{C}}$.

2.1 EMP Algorithm for noiseless input signals

We have the measurement vector \mathbf{y} , the matrix of representation basis $\mathbf{\Psi}$ and the matrix of measurement functions $\mathbf{\Phi}$. They are related by (4). Our

aim as in (13) is achieved by

$$\min_{\hat{\mathbf{C}}} \{H(\mathbf{y}|\mathbf{A})\} \text{ subject to } \|\mathbf{y} - \hat{\mathbf{y}}\|_2 < \varepsilon, \text{ where } \hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{C}}; \varepsilon > 0. \quad (14)$$

In each iteration of the algorithm we extract one component of the signal, that carries maximum information, from the residual \mathbf{e} and refine the approximation $\hat{\mathbf{y}}$. Only one \hat{c}_i changes in the calculation of $\hat{\mathbf{y}}$, but all the M components of \mathbf{e} have the flexibility to change. Hence, more importance has to be given in reducing $H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)})$ compared to the increase in $H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)})$.

Our aim is to capture maximum information of the signal from the residue \mathbf{e} using each coefficient \hat{c}_i and minimize $H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)})$, thus minimizing the conditional entropy $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$. In the absence of noise, the residual \mathbf{e} contains contributions solely from \mathbf{y} and hence \mathbf{e} approaches zero with every iteration. To facilitate the convergence of the algorithm, residual signal energy should decrease in every iteration.

EMP Algorithm for noiseless case

Task

To find a sparse representation \mathbf{C} of a signal \mathbf{S} in Ψ domain subject to $\mathbf{y} = \mathbf{A}\mathbf{C}$, $\|\mathbf{y} - \hat{\mathbf{y}}\|_2 < \varepsilon$. $\mathbf{A} = \Phi\Psi$, where Φ is the measurement matrix, and $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{C}}$ with $\hat{\mathbf{C}}$ as the estimated coefficient vector.

Parameters

Given \mathbf{A} whose columns form a frame, the measurements \mathbf{y} and error threshold ε .

Initialization

- a. Save the norm of \mathbf{y} .
- b. Approximation basis set $\hat{\mathbf{A}}$ with an $M \times N$ matrix of zeros.
- c. Measurement vector approximation $\hat{\mathbf{y}}^{(0)}$ with $M \times 1$ vector of zeros.
- d. Estimated representation $\hat{\mathbf{C}}^{(0)}$ to $N \times 1$ vector of zeros.
- e. Iteration index m to 0.
- f. Residual vector $\mathbf{r}^{(0)}, \mathbf{e}^{(0)}$ to normalized \mathbf{y} .
- g. Magnitude of error vector to unity $\|\mathbf{r}^{(0)}\|_2 = 1$.
- h. Set ε to the given threshold, $w_1 = \frac{N}{N+1}$ and $w_2 = \frac{1}{N+1}$.

Main iteration

Increment m by 1 and perform the following steps.

for each column index $j = 1$ to N

{

$c = \langle \mathbf{r}^{(m-1)}, \mathbf{A}_j \rangle$, \mathbf{A}_j is the j^{th} column of \mathbf{A}

$\hat{\mathbf{y}}_{temp}^{(m)} = \hat{\mathbf{y}}^{(m-1)} + c\mathbf{A}_j$

$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}_{temp}^{(m)}$

Calculate $H(\mathbf{e}|\hat{\mathbf{A}}^{(m)}) = \sum_i e_i^2 \log \frac{1}{e_i^2}$ and

$H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)}) = \sum_i \hat{c}_i^2 \log \frac{1}{\hat{c}_i^2} + c^2 \log \frac{1}{c^2}$ for normalized coefficients

Find the index j_0 which minimizes

$H(\mathbf{y}|\hat{\mathbf{A}}^{(m)}) = (w_1 H(\mathbf{e}|\hat{\mathbf{A}}^{(m)}) + w_2 H(\hat{\mathbf{C}}))$ and $\|\mathbf{e}^{(m)}\|_2 < \|\mathbf{e}^{(m-1)}\|_2$.

}

Update Support, residual signal and representation vector

$\hat{\mathbf{A}}^{(m)} = \hat{\mathbf{A}}^{(m-1)}$ replaced with \mathbf{A}_{j_0} at the j_0^{th} position.

$\mathbf{e}^{(m)} = \mathbf{e}^{(m-1)} - \langle \mathbf{e}^{(m-1)}, \mathbf{A}_{j_0} \rangle \mathbf{A}_{j_0}$

$\hat{c}_{j_0} = \langle \mathbf{e}^{(m-1)}, \mathbf{A}_{j_0} \rangle$

$\hat{\mathbf{C}}^{(m)} = \hat{\mathbf{C}}^{(m-1)}$ added with \hat{c}_{j_0} at the j_0^{th} position.

$\|\mathbf{e}^{(m)}\|_2 = \left(\sum_{i=1}^N |e_i^{(m)}|^2 \right)^{\frac{1}{2}}$

$\mathbf{r}^{(m)} = \mathbf{e}^{(m)}$

Stopping rule

If $\|\mathbf{e}^{(m)}\|_2 < \varepsilon$, stop. Otherwise apply another iteration.

Output

Required sparse representation in $\hat{\mathbf{C}}$. Norm of the signal is restored using the value saved before normalization step. Reconstructed signal $\hat{\mathbf{S}} = \Psi \hat{\mathbf{C}}$.

Entropy reduction need not always minimize the residual signal energy

in each iteration. An iteration could choose a vector \mathbf{A}_i from \mathbf{A} , orthogonal

to the error vector, which will not reduce the error energy. But this situation is avoided in this algorithm by rejecting vectors from \mathbf{A} which do not reduce the residual energy. To this end, the residual signal is projected onto the chosen vector \mathbf{A}_i , and a new component is added to the approximated signal only if the component along the direction of \mathbf{A}_i reduces the residual energy. Otherwise \mathbf{A}_i is discarded. A vector \mathbf{A}_i which minimizes the conditional entropy $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$ as in (12) with reduced error is selected in each iteration. Iterations continue till $\|\mathbf{y} - \hat{\mathbf{y}}\|_2 < \varepsilon$, where ε , the error tolerance permitted by the application, has a value close to zero.

2.2 Convergence of EMP Algorithm

We consider a signal \mathbf{S} with its representation $\mathbf{C} \in \mathbb{R}^N$. A finite dimensional vector space is complete. Since Ψ is a normalized basis in \mathbb{R}^N , \mathbf{S} can be represented as a linear combination of the elements of Ψ with zero residual. Convergence of the algorithm is proved by showing that the signal approximation, $\hat{\mathbf{S}}^{(m)}$, due to the iterations in the algorithm, results in a Cauchy sequence in \mathbb{R}^N . Here we assume that Φ satisfies the restricted isometry property for sparsity $2K$, thus guaranteeing unique recovery of a K sparse signal. Further, we assume that the RIC for sparsity K is very small. Hence the problem of finding \mathbf{S} from \mathbf{y} reduces to the problem of finding \mathbf{C} from \mathbf{S} . Selecting a column from the matrix \mathbf{A} is equivalent to the selection of the corresponding column in Ψ . In the beginning of the first iteration, the residual $\mathbf{r}^{(0)} = \mathbf{S}$. In iteration $(m+1)$, $\mathbf{r}^{(m)} = \mathbf{S} - \hat{\mathbf{S}}^{(m)}$ can be expressed as

$$\mathbf{r}^{(m)} = \left\langle \mathbf{r}^{(m)}, \psi_{i(m+1)} \right\rangle \psi_{i(m+1)} + \mathbf{r}^{(m+1)}, \quad (15)$$

where $\boldsymbol{\psi}_{i(m+1)}$ is the i -th column of $\boldsymbol{\Psi}$ selected in the $(m+1)$ -th iteration, $\hat{\boldsymbol{S}}^{(m)}$ is the approximation of \boldsymbol{S} in the m -th iteration and $\boldsymbol{r}^{(m)}$ is the residual which resulted out of $\hat{\boldsymbol{S}}^{(m)}$.

$$\left| \langle \boldsymbol{r}^{(m)}, \boldsymbol{\psi}_{i(m+1)} \rangle \right|^2 = \left\| \langle \boldsymbol{r}^{(m)}, \boldsymbol{\psi}_{i(m+1)} \rangle \boldsymbol{\psi}_{i(m+1)} \right\|_2^2, \text{ since } \left\| \boldsymbol{\psi}_{i(m+1)} \right\|_2^2 = 1. \quad (16)$$

The residual $\boldsymbol{r}^{(m+1)}$ is orthogonal to $\boldsymbol{\psi}_{i(m+1)}$. Using (15) and (16),

$$\left\| \boldsymbol{r}^{(m)} \right\|_2^2 = \left| \langle \boldsymbol{r}^{(m)}, \boldsymbol{\psi}_{i(m+1)} \rangle \right|^2 + \left\| \boldsymbol{r}^{(m+1)} \right\|_2^2. \quad (17)$$

Using (15) to (17),

$$\left\| \boldsymbol{S} \right\|_2^2 = \left\| \boldsymbol{r}^{(0)} \right\|_2^2 = \sum_{m=0}^N \left| \langle \boldsymbol{r}^{(m)}, \boldsymbol{\psi}_{i(m+1)} \rangle \right|^2 + \left\| \boldsymbol{r}^{(N+1)} \right\|_2^2. \quad (18)$$

From (17),

$$\left\| \boldsymbol{r}^{(m+1)} \right\|_2^2 = \left\| \boldsymbol{r}^{(m)} \right\|_2^2 - \left| \langle \boldsymbol{r}^{(m)}, \boldsymbol{\psi}_{i(m+1)} \rangle \right|^2. \quad (19)$$

Hence, $\frac{\left\| \boldsymbol{r}^{(m+1)} \right\|_2^2}{\left\| \boldsymbol{r}^{(m)} \right\|_2^2} = 1 - \frac{\left| \langle \boldsymbol{r}^{(m)}, \boldsymbol{\psi}_{i(m+1)} \rangle \right|^2}{\left\| \boldsymbol{r}^{(m)} \right\|_2^2} \leq 1$. Equality arises when the chosen vector is orthogonal to $\boldsymbol{r}^{(m)}$. Orthogonal vectors are discarded in the algorithm as they will not refine the residual signal. Thus, $\{\boldsymbol{r}^{(m)}; m = 1, 2, \dots\}$ is a bounded decreasing sequence making $\{\hat{\boldsymbol{S}}^{(m)}; m = 1, 2, \dots\}$ a bounded increasing sequence bounded above at $\left\| \boldsymbol{S} \right\|_2$. Since $\left\| \hat{\boldsymbol{S}}^{(m)} \right\|_2$ is not the upper bound, there exists an integer L such that

$$\left\| \boldsymbol{S} - \hat{\boldsymbol{S}}^{(m)} \right\|_2 \leq \left\| \boldsymbol{S} - \hat{\boldsymbol{S}}^{(L)} \right\|_2, \text{ for } m \geq L. \quad (20)$$

Thus,

$$\left\| \mathbf{S} - \hat{\mathbf{S}}^{(m+1)} \right\|_2 \leq \left\| \mathbf{S} - \hat{\mathbf{S}}^{(m)} \right\|_2 \leq \left\| \mathbf{S} - \hat{\mathbf{S}}^{(L)} \right\|_2, \text{ for } m \geq L. \quad (21)$$

Hence $\{\hat{\mathbf{S}}^{(m)}, m = 1, 2, \dots\}$ is Cauchy. Equivalently, $\{\|\Psi(\mathbf{C} - \hat{\mathbf{C}}^{(m)})\|_2; m = 1, 2, \dots\}$ is a bounded decreasing sequence which implies that the set of representation coefficients, $\{\hat{\mathbf{C}}^{(m)}; m = 1, 2, \dots\}$, forms a Cauchy sequence.

The EMP algorithm may not give the exact representation with the sparsest set of coefficients in a finite number of iterations unless the vectors are orthonormal. In the noise-free case, it is possible to recover the sparse set of coefficients ideally, as convergence of the algorithm is guaranteed. The algorithm can be modified to calculate the vector of sparse representation coefficients, $\hat{\mathbf{C}}$, in a finite number of iterations. This calculation is based on the minimization of $\|\mathbf{y} - \hat{\mathbf{y}}\|^2$, after choosing a candidate vector from \mathbf{A} by minimizing $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$, in each iteration. This can lead to recovery in K iterations, where K is the sparsity of the signal, as guaranteed by the OMP algorithm [7].

2.3 EMP Algorithm for noisy input signals

When the input signal is noisy, the aim of EMP algorithm in each iteration is to reduce the overall conditional entropy calculated as in (12). We assume white noise having dense representation $\tilde{\mathbf{C}}$ relative to the basis Ψ in which the signal has a sparse representation \mathbf{C} . The noisy signal \mathbf{S} is

$$\mathbf{S} = \mathbf{S}_o + \mathbf{S}_n, \quad (22)$$

where \mathbf{S}_o and \mathbf{S}_n are the signal and noise components, respectively. The corresponding noisy measurement vector is

$$\mathbf{y} = \mathbf{y}_o + \mathbf{y}_n, \quad (23)$$

where \mathbf{y}_o and \mathbf{y}_n are the signal and noise components, respectively, in the measurement vector. Using the notions of representation and measurement,

$$\mathbf{y}_o = \Phi \mathbf{S}_o = \Phi \Psi \mathbf{C} \quad (24)$$

$$\mathbf{y}_n = \Phi \mathbf{S}_n = \Phi \Psi \tilde{\mathbf{C}} \quad (25)$$

$$\mathbf{y} = \Phi \Psi (\mathbf{C} + \tilde{\mathbf{C}}) = \mathbf{A}(\mathbf{C} + \tilde{\mathbf{C}}), \quad (26)$$

where $\mathbf{A} = \Phi \Psi$ is of full row rank, and hence its pseudo-inverse exists. The vector \mathbf{C} is sparse but $\tilde{\mathbf{C}}$ is dense as the representation basis represents signal sparsely and noise densely. If \mathcal{R} represents a recovery algorithm, then $\mathcal{R}(\mathbf{y}) = \mathbf{C}$. Conventional recovery algorithms work on least square error minimization and hence are not capable of distinguishing between \mathbf{C} and $\tilde{\mathbf{C}}$. The only distinguishing factor between \mathbf{C} and $\tilde{\mathbf{C}}$ is that \mathbf{C} is sparse and $\tilde{\mathbf{C}}$ is dense relative to the chosen representation basis Ψ . EMP algorithm makes use of entropy minimization method which has the inherent capability to distinguish and extract sparse coefficients. The component $\tilde{\mathbf{C}}$ in the sum $\mathbf{C} + \tilde{\mathbf{C}}$ increases the entropy of representation. In the proposed EMP algorithm, noise components are rejected such that the increase in conditional entropy of signal representation is restricted below a predefined limit γ , in each iteration of the algorithm. When \mathbf{C} and $\tilde{\mathbf{C}}$ are normalized,

$$H(\mathbf{C}) < H(\tilde{\mathbf{C}}), \quad (27)$$

where $H(\mathbf{C})$ and $H(\tilde{\mathbf{C}})$ are respectively, the entropy of \mathbf{C} and $\tilde{\mathbf{C}}$ calculated according to (7). The morphological component separation presented in [17] can be used to separate noise and signal components based on their respective sparsity in a given basis. The goal of the EMP algorithm under noisy conditions is

$$\min_{\hat{\mathbf{C}}} H(\mathbf{y}|\mathbf{A}) \text{ subject to } \|\mathbf{y} - \hat{\mathbf{y}}\|_2 < \varepsilon, \text{ and } \Delta H(\mathbf{y}) < \gamma, \quad (28)$$

where $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{C}}$, $\Delta H(\mathbf{y}) = H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})/H(\mathbf{y}|\hat{\mathbf{A}}^{(m-1)})$, $\gamma > 0$, $\varepsilon > 0$.

The algorithm starts with the measured signal \mathbf{y} as residual \mathbf{e} , having a maximum of M nonzero elements, and the estimated coefficients $\hat{\mathbf{C}}$ as a zero vector having N elements. The measurement \mathbf{y} which is the sum of its approximation and residual, is thus represented using M coefficients out of the overall $N + M$ coefficients. Initial iterations result in significant non-zero coefficients in $\hat{\mathbf{C}}$, with each nonzero coefficient capturing the \mathbf{y}_o component from \mathbf{y} . In iteration m ,

$$\mathbf{y} = \hat{\mathbf{y}}^{(m)} + \mathbf{e}^{(m)}, \quad (29)$$

where

$$\hat{\mathbf{y}}^{(m)} = \sum_i^m \mathbf{A}_i \hat{\mathbf{C}}_i. \quad (30)$$

The entropy $H(\hat{\mathbf{C}})$ increases in every iteration as a non-zero component gets added to $\hat{\mathbf{C}}$. Since EMP algorithm works on minimizing conditional entropy, the column of matrix \mathbf{A} that induces sparsity in the resulting residue is chosen in each iteration. The conditional entropy $H(\{\mathbf{y} - \hat{\mathbf{y}}\}|\hat{\mathbf{A}}^{(m)})$ of error, given the measurement vector and its approximation, decreases as components of signal present in the residue are transferred to $\hat{\mathbf{y}}$ and captured in $\hat{\mathbf{C}}$. The energy of the residual decreases considerably. Initial iterations result in

overall decrease or at the most marginal increase in $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$. Removal of the \mathbf{S}_o components from \mathbf{S} causes removal of information about the signal from the residue, thus minimizing its conditional entropy $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$. Conditional entropy of \mathbf{y} is reduced significantly by representing its \mathbf{y}_o component using sparse $\hat{\mathbf{C}}$. Therefore, initial iterations capture \mathbf{S}_o in approximating \mathbf{y} by $\hat{\mathbf{y}}$.

The components of noise along the basis chosen in each iteration are also present in the representation. Subsequent iterations have significant noise in the residue \mathbf{y}_n due to \mathbf{S}_n which cannot be sparsely represented by $\hat{\mathbf{C}}$, relative to Ψ . Therefore, reduction in the first term in (12) is considerably lower than the increase in the second term which effectively leads to increase in $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$. In the absence of noise, the energy of the residual signal would not exceed ε , and $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)}) \approx H(\hat{\mathbf{y}}|\hat{\mathbf{A}}^{(m)})$ when the algorithm converges.

When the input is noisy, prior to choosing a particular vector for refining the representation, an additional step in the algorithm compares $H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})$ and $H(\mathbf{y}|\hat{\mathbf{A}}^{(m-1)})$ which are the conditional entropy of the measurement given the chosen subset of vectors in $\hat{\mathbf{A}}$, in iteration m and $(m-1)$, respectively. The algorithm proceeds to update only if

$\Delta H(\mathbf{y}) = H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})/H(\mathbf{y}|\hat{\mathbf{A}}^{(m-1)})$ is less than γ . The parameter γ decides whether a new component is to be added to the signal approximation at the cost of increase in the conditional entropy. It increases linearly with the SNR of input signal allowing more components to be added to the representation. In the noiseless case, γ is infinite which reduces residual energy even if it causes increase in the conditional entropy between successive iterations as in (14). In the noisy case, components which cause increase in the conditional entropy beyond γ is attributed to noise components and are rejected by the algorithm. Iterations terminate when either the increment in the conditional

entropy between consecutive iterations is beyond the permitted limit γ or, the error energy threshold requirement is met.

EMP Algorithm for noisy case

Initialization

Set $w_1 = \frac{M - \|\hat{\mathbf{C}}\|_0}{M}$ and $w_2 = \frac{\|\hat{\mathbf{C}}\|_0}{M}$.

(Other steps are the same as in the algorithm for the case of noiseless measurements)

Main iteration

for each column index $j = 1$ to N

{

(steps are the same as in the algorithm for the case of noiseless measurements)

}

$$\Delta H(\mathbf{y}) = \frac{H(\mathbf{y}|\hat{\mathbf{A}}^{(m)})}{H(\mathbf{y}|\hat{\mathbf{A}}^{(m-1)})}.$$

Stop the iterations if $\Delta H(\mathbf{y}) < \gamma$, else proceed to next section.

Update Support, residual signal and representation vector

(remaining steps are the same as in the algorithm for the case of noiseless measurements)

3 Results and Discussion

The scheme for compressed sensing which uses the EMP algorithm for recovery is studied using both synthetic test signals and actual speech signals. Speech signal is used as an example of a compressible signal. Performance of the scheme under both noisy and noiseless conditions is evaluated. Recovery percentage is used for performance evaluation under noise-free conditions when sparsity is known. Signal to Reconstruction Error Ratio (SRER) is

used as an objective measure of performance of the scheme under noiseless condition when sparsity is not known. It gives a measure of capability of the algorithm to approximate the input signal \mathbf{S} from the measurements \mathbf{y} faithfully. Signal to Noise Ratio (SNR) is used for evaluating the performance under noisy environment. SNR measures the ability of the algorithm to reject noise in the input signal \mathbf{S} . When the algorithm has ability to reject noise, reconstruction from measurements from a noisy signal \mathbf{S} will result in low SRER and high SNR. Since the EMP algorithm aims to minimize conditional entropy and not energy of the residual, we use information power as another parameter for performance comparison. Information power (IP) is defined as the variance of a Gaussian source required to make its entropy equal to that of the entropy $H(\mathbf{X})$ of the signal source \mathbf{X} . *i.e*; $H(\mathbf{G}) = H(\mathbf{X})$, where $H(\mathbf{G})$ is the entropy of the Gaussian source. Reduction in IP indicates sparse or compressible representation. SRER is calculated as

$$SRER = 10 \log_{10} \frac{\sum_i S_i^2}{\sum_i (S_i - \hat{S}_i)^2}, \quad (31)$$

where S_i represents samples of input signal and \hat{S}_i represents samples reconstructed from the recovered sparse set of coefficients. SNR is calculated as

$$SNR = 10 \log_{10} \frac{\sum_i S_i^2}{\sum_i (S_i - \hat{S}_i)^2}, \quad (32)$$

where S_i represents samples of input signal before adding noise and \hat{S}_i represents samples reconstructed from the sparse estimate of the representation coefficients based on measurements from noisy signal. The entropy of a Gaussian source with variance σ is $H(\mathbf{G}) = \log_b \sqrt{2\pi e \sigma^2}$, where b is the

base to be considered. Information power is

$$\sigma^2 = \frac{2^{2H(\mathbf{X}) \log_2 b}}{2\pi e}. \quad (33)$$

Simulations were done with noisy signals generated by adding white Gaussian noise with input SNR varying from $-6dB$ to $3dB$.

3.1 Experiment with synthetic signal

Simulations were carried out to study the performance of various recovery algorithms under the following cases.

1. Sparse input signal with orthogonal representation basis and known sparsity.
2. Sparse input signal with orthogonal representation basis and unknown sparsity.
3. Sparse input signal with non-orthogonal representation basis added with $3dB$ white Gaussian noise and unknown sparsity.
4. Compressible input signal with non-orthogonal representation basis added with $3dB$ white Gaussian noise.

Synthetic test signals are generated by linear combination of representation basis vectors. Fourier representation basis was used as orthogonal basis for the study of case-1. The sparsity of the signal considered was $K = 4$. Fig. 1 shows that the performance of EMP algorithm is at par with other recovery algorithms in the absence of noise.

Fig. 2 shows the performance of various algorithms when sparsity is not known in advance. In this study also, a signal of sparsity $K = 4$ was used. But the algorithms were not presented with the information about sparsity.

Conventional greedy algorithms estimate sparsity from the measurements. The results show that EMP performs marginally better than the other algorithms when number of measurements is low.

The performance of various algorithms in the presence of $3dB$ noise is presented in Table 1. In this study, sparsity of the signal is considered unknown and non-orthogonal representation basis is used. The experiment was carried out for signals having dimension 40 and sparsity $K = 4$; the sparsity was not an input to algorithms. The results show that at low SNR, OMP performs better than all other conventional greedy algorithms. EMP outperforms OMP for reduced number of measurements and is marginally better at increased number of measurements. Estimated sparsity was used for halting OMP, CoSaMP and ROMP. The estimated sparsity varies from 3 to 5 corresponding to measurements varying from 20 to 36. It is observed that as the algorithms approximate the signal with more components beyond its sparsity, more noise gets into the signal approximation thus causing reduction in SNR of the reconstructed signal. Since greedy pursuits estimate sparsity from the number of measurements, increase in the number of measurements leads to increase in the estimated sparsity, resulting in decrease of SNR. This is avoided in EMP as the algorithm does not iterate using sparsity as a parameter.

The performance of various algorithms for a compressible signal in the presence of $3dB$ noise is presented in Table 2. Since the signal is not strictly sparse, the halting condition of conventional algorithms is changed such that the algorithms terminate when residual norm is below the predetermined threshold. The results indicate that performance of EMP is superior to all other greedy pursuits when the signal is compressible. SNR was averaged over 50 runs with different noisy inputs for studying the performance of the

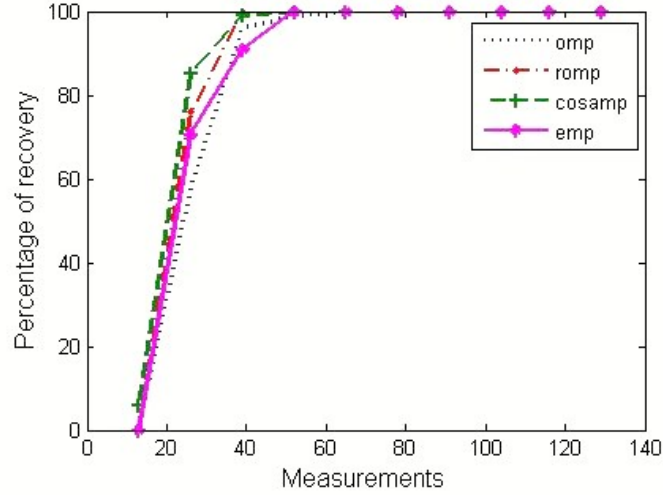


Figure 1: Performance comparison of various recovery algorithms on a 4-sparse signal of dimension 200 under noiseless conditions.

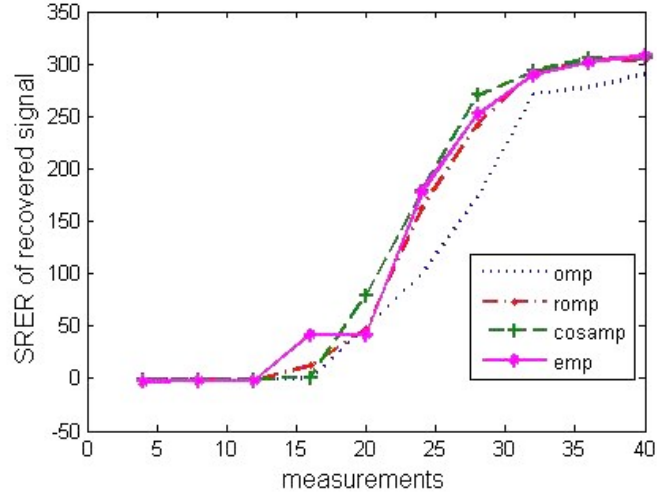


Figure 2: Performance comparison of various recovery algorithms on a 4-sparse signal of dimension 200 under noiseless conditions, with unknown sparsity.

algorithms under noisy conditions.

Table 1: Comparison of SNR for synthetic sparse signal of length 40 with sparsity unknown, recovered using various greedy pursuits at 3dB noise

No. of measurements	Reconstruction SNR in dB			
	EMP	OMP	CoSaMP	ROMP
20	0.77	0.02	-0.59	0.35
24	1.74	0.40	-0.02	1.43
28	2.57	1.42	-0.01	1.45
32	3.21	3.10	1.87	1.55
36	3.61	3.27	-0.79	0.87

Table 2: Comparison of SNR for synthetic compressible signal of length 40, recovered using various greedy pursuits at 3dB noise

No. of measurements	Reconstruction SNR in dB			
	EMP	OMP	CoSaMP	ROMP
20	2.33	2.07	1.14	-0.07
24	2.65	2.83	2.42	0.86
28	2.93	2.70	1.53	-0.83
32	2.88	2.84	1.41	-0.77
36	3.21	3.03	-0.21	-0.62

Table 3: Comparison of IP for compressible signal of dimension 40 at 0dB noise recovered using various greedy pursuits with 36 measurements.

Algorithm	IP	SNR in dB
EMP	3.06	3.94
OMP	9.53	1.87
CoSaMP	61.28	1.10
ROMP	50.32	0.98

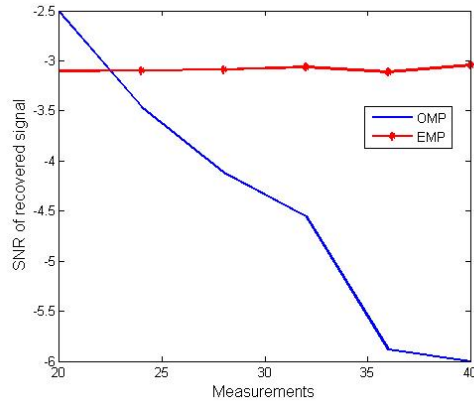
When the sparsity is unknown, it is estimated from the number of measurements, M , and used in the algorithms as $K = M/(2\log_e N)$, where N is the signal dimension. When non-orthogonal representation basis is considered, the performance of greedy pursuits that update more than one component in an iteration is poor in comparison with the OMP algorithm which chooses only one component per iteration.

As entropy decreases, information power also decreases. Table 3 shows information power of a noisy signal at 0 dB reconstructed using algorithms under consideration. Information power of the original signal without noise was 2.17 and information power of noisy signal was 20.51. The base b considered for logarithm in (33) was 2. Information power clearly indicates the ability of EMP to reject noise components, that require dense representation and thus have high entropy.

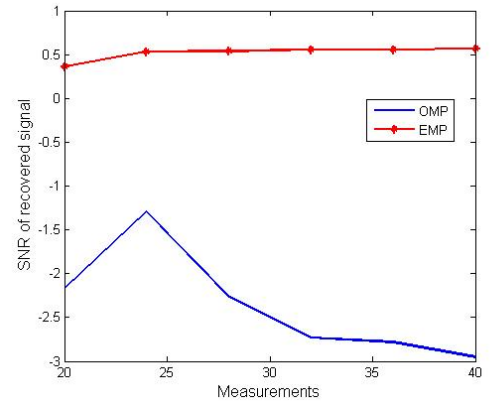
Fig. 3 shows comparison of the performances of OMP and EMP algorithms for various levels of input noise. Non-orthogonal representation basis was used to generate the synthetic signal. Simulation study shows that performance of EMP at SNR ranging from $-6dB$ to $3dB$ is higher than that of all other algorithms considered, indicating the capability of the EMP algorithm for providing robustness to the CS system in the presence of noise.

3.2 Experiment with speech segment

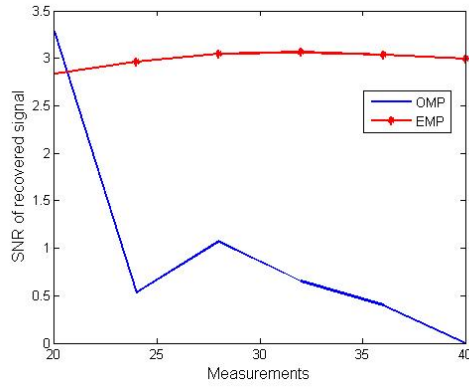
This experiment makes use of the representation basis obtained in [15]. A speech segment sampled at Nyquist rate is used as input. Wavelet packet basis is used as the representation basis. Wavelet packet basis obtained using the best basis algorithm proposed in [14] is chosen to arrive at the tree structure suitable for speech signals [15]. Impulse response at each of the terminal nodes of the tree and their translates constituted the representation basis. The details of the signal representation can be found in [15]. The representation basis is non-orthogonal. Measurement matrix is chosen based on the Multi-Dimensional Scaling method proposed in [18]. This experiment benchmarks the performances of the algorithms when the signal is compressible in a non-orthogonal basis with its sparsity unknown. Perceptual quality of the reconstructed signal was indistinguishable from the



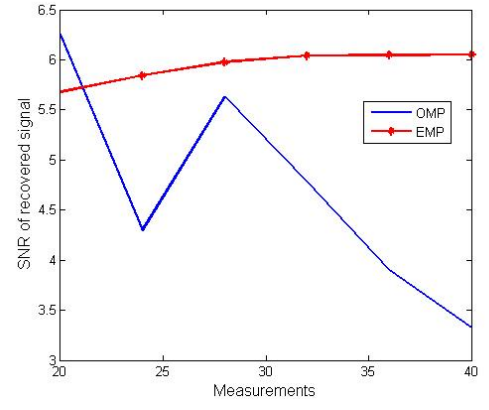
(a) Input SNR = -6dB



(b) Input SNR = -3dB



(c) Input SNR = 0dB



(d) Input SNR = 3dB

Figure 3: Performance comparison of EMP and OMP as recovery algorithms on a noisy synthetic signal at various noise levels.

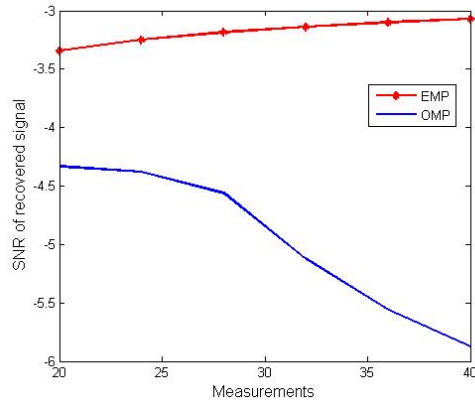
original in the noise-less case for the number of measurements as low as 60% of that required at Nyquist rate. Table 4 shows that the performance of the EMP algorithm is at par with the OMP algorithm under noise-less case.

Table 4: Comparison of SRER (in dB) for speech signal divided into segments of length 40, recovered through the OMP and the EMP algorithms.

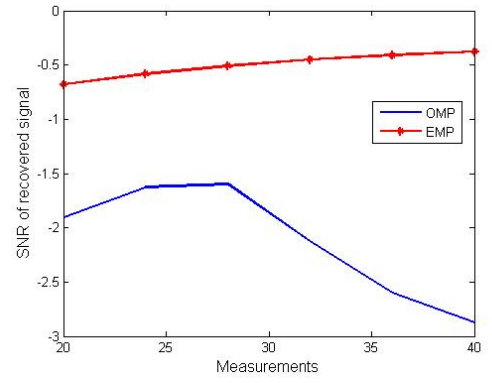
No. of measurements	OMP	EMP
16	2.93	3.30
20	7.77	8.03
24	11.32	11.76
28	13.44	13.77
32	17.96	18.03
38	26.83	27.30
40	267.74	267.60

Fig. 4 shows the performance when speech signal with noise is compressively measured and reconstructed using the EMP and the OMP algorithms. The halting condition of the OMP algorithm was based on residual error threshold and not based on the estimated sparsity since the input signal was compressible and not exactly sparse. The figure shows the superior performance of the EMP algorithm.

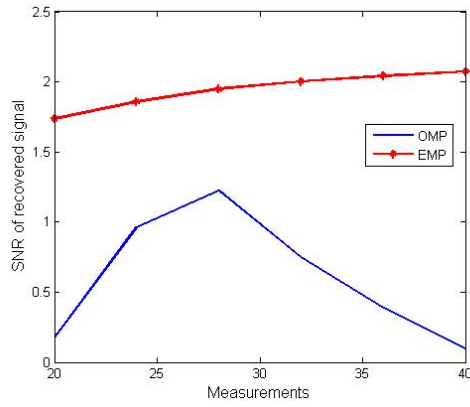
The simulation results show that the EMP algorithm has the capability to perform at par with the CoSaMP, ROMP and OMP algorithms when the signal is sparse and without any noise. But, for noisy compressible signals, the EMP algorithm performs significantly better than the other greedy algorithms mentioned. With low SNR, reducing number of measurements leads to smaller estimated sparsity which improves the performance of the OMP, CoSaMP and ROMP algorithms. Since the EMP algorithm tries to reduce conditional entropy, it has the ability to reject noise that requires dense coefficients in the chosen representation basis for the signal. The performance is more or less independent of the number of measurements. But



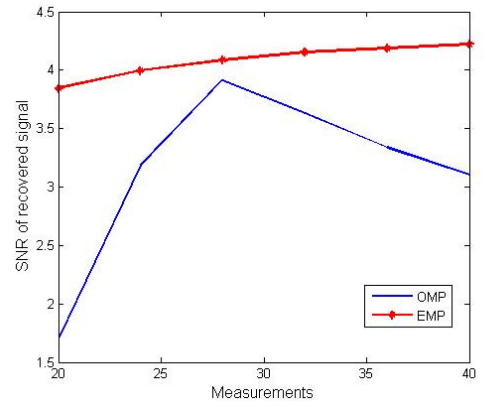
(a) Input SNR = -6dB



(b) Input SNR = -3dB



(c) Input SNR = 0dB



(d) Input SNR = 3dB

Figure 4: Performance comparison of the EMP and the OMP recovery algorithms on a noisy speech signal at various noise levels.

computational overhead of the EMP algorithm is higher than that for the other algorithms as entropy calculation is computationally intensive. In the EMP algorithm, the chosen threshold value ε has to meet the constraints imposed by applications on the required minimum SRER under noiseless case. When noisy signal is presented, the relaxation parameter γ should be fine-tuned in accordance with the SNR of the input signal. In our study, γ was calculated as $(M + N + 5SNR)/M$, where SNR is in dB.

4 CONCLUSION

In this article, we have presented the EMP algorithm for sparse signal recovery under noiseless and noisy conditions. The sparse recovery scheme based on the proposed EMP algorithm is unique in its robustness in the presence of noise. The classical greedy algorithms reconstruct the original signal faithfully but fail to separate noise. The functionality of the EMP algorithm is based on conditional entropy minimization instead of energy minimization which facilitates its noise resilience. The EMP Algorithm does not require the sparsity to be known, at the same time, it offers significant improvement in the SNR of the reconstructed signal in the presence of noise. Noise components do not get added to the reconstructed signal even when measurements are increased. Therefore, the EMP algorithm may replace the conventional greedy pursuit algorithms when the measurements are noisy. Under noiseless conditions, the performance of the EMP algorithm is at par with the best performance of other greedy algorithms.

References

- [1] E. J. Candés and M. B. Wakin, “An Introduction To Compressive Sampling,” *IEEE SP Mag.*, Vol. 25, No. 2, pp. 21–30, March 2008.
- [2] R. G. Baraniuk, “Compressive Sensing [Lecture Notes],” *IEEE SP Mag.*, Vol. 24, No. 4, pp. 118–120, July 2007.
- [3] E. J. Candés, T. Tao, “Decoding By Linear Programming,” *IEEE Trans. on Inf. Th.*, Vol. 51, No. 12, pp. 4203–4215, Dec. 2005.
- [4] G. Peyre, “Best Basis Compressed Sensing,” *IEEE Trans. on SP*, Vol. 58, No. 5, pp. 2613–2622, May 2010.
- [5] S. Ravishankar, Y. Bresler, “Learning Sparsifying Transforms,” *IEEE Trans. on SP*, Vol. 61, No. 5, pp. 1072–1086, March 2013.
- [6] S. G. Mallat, “A Wavelet Tour of Signal Processing,” Academic Pr, 1999.
- [7] J. A. Tropp and A. C. Gilbert, “Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit,” *IEEE Trans. on Inf. Th.*, Vol. 53, No. 12, pp. 4655–4666, Dec. 2007.
- [8] J. Wang, S. Kwon, and B. Shim, “Generalized Orthogonal Matching Pursuit,” *IEEE Trans. on SP*, Vol. 60, No. 12, pp. 6202–6216, Dec. 2012.
- [9] B. Mailhé, R. Gribonval, P. Vandergheynst, F. Bimbot, “Fast Orthogonal Sparse Approximation Algorithms Over Local Dictionaries,” *Signal Processing*, Vol. 91, No. 12, pp. 2822–2835, Jan. 2011.
- [10] D. Needell and J. A. Tropp, “CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples,” *Appl. Comp. Harmon. Anal.*, Vol. 26, No. 3, pp. 301–321, May. 2009.

- [11] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, “Sparse Solution Of Underdetermined Systems Of Linear Equations By Stagewise Orthogonal Matching Pursuit,” *IEEE Trans. on Inf. Th.*, Vol. 58, No. 2, pp. 1094–1121, Feb. 2012.
- [12] D. Needell and R. Vershynin, “Signal Recovery from Incomplete and Inaccurate Measurements via ROMP,” *IEEE J. Sel. Topics in Signal Proc.*, Vol. 4, No. 2, pp. 310–316, Apr. 2010.
- [13] M. A. Davenport, P. T. Boufounos, and R. G. Baraniuk, “Compressive Domain Interference Cancellation,” Author manuscript, published in SPARS’09 - Signal Processing with Adaptive Sparse Structured Representations (2009).
- [14] R. R. Coifman and M. V. Wickerhauser, “Entropy Based Algorithms for Best Basis Selection,” *IEEE Trans. on Inf. Th.*, Vol. 38, No. 2, pp. 713–718, March 1992.
- [15] V. Meena, G. Abhilash, “Sparse Representation and Recovery of a Class of Signals Using Information Theoretic Measures,” INDICON, 2013. DOI:10.1109/INDCON.2013.6725897
- [16] J. C. A. van der Lubbe, *Information Theory*, Camb. Univ. Press, 1997.
- [17] J. Bobin, Y. Moudden, J. L. Starck and M. Elad, “Morphological Diversity and Source Separation,” *IEEE SP Letters*, Vol. 13, No. 7, pp. 409–412, July 2006.
- [18] A. Ghodsi, “Dimensionality Reduction, A Short Tutorial, Technical Report,” 2006-14, Dept. of Statistics and Actuarial Science, Univ. of Waterloo, 2006.

- [19] S. Stankovic, L.J. Stankovic, and I. Orovic, “A Relationship between the Robust Statistics Theory and Sparse Compressive Sensed Signals Reconstruction,” IET Signal Processing, 2014